

**PROVIDING SPEAKER IDENTIFYING INFORMATION WITHIN  
EMBEDDED DIGITAL INFORMATION**

**Inventor(s):**

Thomas E. Creamer  
Peeyush Jaiswal  
Victor S. Moore

**International Business Machines Corporation**

IBM Docket No. BOC9-2003-0094  
IBM Disclosure No. BOC8-2003-0120

## **PROVIDING SPEAKER IDENTIFYING INFORMATION WITHIN EMBEDDED DIGITAL INFORMATION**

### **BACKGROUND**

#### **Field of the Invention**

[0001] The invention relates to speaker identification over a communications channel.

#### **Description of the Related Art**

[0002] Conventional calling line identification (CLID) and the associated display terminals have become commonplace in the market. Known CLID services deliver the directory number, subscriber name or business name associated with the calling telephone line rather than the true identity of the caller. Human recognition of the caller, if known to the called party, must be relied on for verifying a caller's identity. The value of human recognition, however, is limited by the fact that the caller may not be known to the called party. Thus known CLID services fail to provide an assured identity of the caller that can be acted on reliably.

[0003] Consequently, the CLID cannot be acted on with certainty since the same CLID is delivered regardless of who actually places the call. For example, when all members of a household share the same CLID associated with a subscriber number, the displayed name and number does not identify which of the several family members is making the call. If a call is placed by an individual away from their customary phone as would occur for a business traveler at a payphone, hotel room, or colleague's desk, the caller's personal identity is not delivered.

[0004] CLID information is transmitted on the subscriber loop using frequency shift keyed (FSK) modem tones. These FSK modem tones are used to transmit the display message in American Standard Code for Information Interchange (ASCII) character code form. The transmission of the display message takes place between the first and second ring. Hence, the CLID data is not sent once the call is established.

[0005] As such, the aforementioned problems with CLID are further exacerbated within the context of conference calls. With respect to conference calls, once each

participant is connected, it can be difficult for a listener to discern the identity of a speaking party. This may result from the listener's unfamiliarity with the speaker or that several of the conference call participants sound alike. Because CLID is transmitted prior to the telephone call, CLID is not well suited to address this problem.

### **SUMMARY OF THE INVENTION**

**[0006]** The present invention provides a method, system, and apparatus for providing identifying information as well as authentication information to a subscriber. In particular, a call participant, whether prior to a telephone call being established or during a telephone call, can provide some sort of identifier or code. That code can be used to verify the identity of the call participant, or authenticate the call participant. Information specifying whether the person was authenticated can be provided to a subscriber.

**[0007]** One aspect of the present invention can include a method of providing identifying information over a voice communications link. The method can include receiving, from a call participant, a personal identification code over the voice communications link, determining identifying information for the call participant using the personal identification code, and encoding the identifying information of the call participant within a voice stream carried by the voice communications link. Accordingly, the voice stream and identifying information can be sent to a subscriber.

**[0008]** In one embodiment of the present invention, the voice communications link can be a telephony communications link. The identifying information can indicate whether the call participant has been authenticated. The identifying information and the voice stream can be digital information, such that the identifying information is embedded within the voice stream. For example, the encoding step can include removing inaudible portions a speech signal and embedding the identifying information in place of the inaudible portions of the speech signal within the voice stream.

**[0009]** The method also can include receiving the voice stream and identifying information and decoding the identifying information. A representation of the identifying information can be presented, for example to the subscriber. An audible representation of the voice stream also can be played. In one embodiment, the audible representation of the received voice stream can be played substantially concurrently with the presentation of the identifying information.

**[0010]** Other embodiments of the present invention can include a system having means for performing the various steps disclosed herein and a machine readable storage for causing a machine to perform the steps described herein.

**BRIEF DESCRIPTION OF THE DRAWINGS**

[0011] There are shown in the drawings, embodiments which are presently preferred, it being understood, however, that the invention is not limited to the precise arrangements and instrumentalities shown.

[0012] FIG. 1 is a schematic diagram illustrating a system for providing speaker identifying information within embedded digital information in accordance with the inventive arrangements disclosed herein.

[0013] FIG. 2 is a flow chart illustrating a method of providing speaker identifying information within embedded digital information in accordance with the inventive arrangements disclosed herein.

### **DETAILED DESCRIPTION OF THE INVENTION**

**[0014]** FIG. 1 is a schematic diagram illustrating a system 100 for providing speaker identifying information within embedded digital information in accordance with the inventive arrangements disclosed herein. As shown, the system 100 can include an Identification and Authentication Service (IAS) 105 and a communications network 110 over which a subscriber 115 and a call participant 120 can communicate. The communications network can include, but is not limited to, the Internet, a wide area network, a local area network, an intranet, and/or the Public Switched Telephone Network.

**[0015]** The IAS 105 can be implemented as a computer program executing within an information processing system. For example, in one embodiment, the IAS 105 can execute within a computer system such as a server that is communicatively linked to a telephony switching system via a suitable gateway interface. In that case, the IAS 105 can be located on premises with the telephony switching system or remote from such a switching system. In another embodiment, the IAS 105 can execute within the telephony switch itself.

**[0016]** The IAS 105 can be configured to join in telephone calls, either prior to the establishment of the telephone call or during the call, to authenticate a call participant 120. A subscriber 115 to the IAS 105 can selectively engage the IAS 105 to authenticate calling parties, such as call participant 120. For example, the IAS 105 can be invoked for particular calls to the subscriber 115. Determinations as to when the IAS 105 is to be invoked can be based upon rules defined by the subscriber 115 which specify dates and times, such as after 9:00 p.m., on national holidays, based upon whether the calling number is recognized, or a combination thereof. Still, the IAS 105 can be invoked for all calls to the subscriber 115, or can be invoked by the subscriber 115 as desired to verify any call participant whether the call participant initiated the call or not.

**[0017]** The IAS 105 can receive identifying information from a call participant 120 and compare that information with stored authentication data. The authentication data can be stored within the IAS 105 or can be stored in a remote data store that is communicatively linked with the IAS 105. In any case, based upon a comparison of

identifying information provided by the call participant 120 with stored authentication information, the IAS 105 can authenticate the call participant 120 to determine whether the call participant 120 is who that person portends to be.

**[0018]** The IAS 105 can encode the identifying information using any of a variety of different mechanisms. For example, in one aspect, the IAS 105 can be implemented as a perceptual audio processor, similar to a perceptual codec, to analyze a received voice signal. A perceptual codec is a mathematical description of the limitations of the human auditory system and, therefore, human auditory perception. Examples of perceptual codecs can include, but are not limited to MPEG Layer-3 codecs and MPEG Layer-4 codecs. The IAS 105 is substantially similar to the perceptual codec with the noted exception that the IAS 105 can, but need not implement, a second stage of compression as is typical with perceptual codecs.

**[0019]** The IAS 105, similar to a perceptual codec, can include a psychoacoustic model to which source material, in this case a voice signal from a call participant or speaker, can be compared. By comparing the voice signal with the stored psychoacoustic model, the perceptual codec identifies portions of the voice signal that are not likely, or are less likely to be perceived by a listener. These portions are referred to as being inaudible. Typically a perceptual codec removes such portions of the source material prior to encoding, as can the IAS 105. The IAS 105, however, can add the identifying information in place of the removed source material, i.e. the inaudible portions of speech.

**[0020]** Still, those skilled in the art will recognize that the present invention can utilize any suitable means or techniques for encoding identifying information and embedding such digital information within a digital voice stream. As such, the present invention is not limited to the use of one particular encoding scheme.

**[0021]** FIG. 2 is a flow chart illustrating a method 200 of providing speaker identifying information within embedded digital information in accordance with the inventive arrangements disclosed herein. The method 200 can begin in a state where a call participant is attempting to place a telephone call to a subscriber or a state where a telephone call has been established between a call participant and a subscriber. The telephone call can be a conventional landline telephone call, a wireless or mobile

telephone call, or a Voice-Over Internet Protocol (VOIP) telephone call.

**[0022]** In step 205, the IAS can prompt the call participant for a personal identification code. The call participant can provide a personal identification code in step 210. The personal identification code can be provided as a series of one or more dual tone multi-frequency (DTMF) tones or as speech. Accordingly, the IAS can interpret the received personal identification code, whether by identifying the keys that were activated in the case of a DTMF input, or by recognizing user speech to determine a text equivalent of the received input.

**[0023]** In step 215, the IAS compares the received personal identification code with stored authentication information. In one embodiment, the IAS can determine a set of authentication information to which the received personal identification code is to be compared by determining the telephone number corresponding to the communications link over which the call participant has provided the personal identification code. For example, each telephone number can be associated with one or more user profiles. Each user profile can be associated with one of several family members or other likely users of the telephone line or number and each specifying a unique personal identification code. In another embodiment, the IAS can first query the call participant for an identifier. The identifier can be used to locate stored authentication information. In any case, having located authentication information for the call participant, the IAS can compare the received personal identification code with a personal identification code stored within the authentication information.

**[0024]** In step 220, the IAS determines whether the call participant has been authenticated based upon the comparison. More particularly, if the received personal identification code matches the code stored within the authentication data, the call participant has been successfully authenticated. In other words, an identity of the call participant has been established and verified. Notably, the identity of the call participant may not be the name associated with the line or telephone number over which that call participant is communicating.

**[0025]** In step 225, the IAS can encode identifying information for the call participant within a voice stream. That is, the identifying information, in digital form, can be embedded within the digital voice stream of the telephone call, resulting in a voice



signal having embedded digital identifying information for the speaker. In one embodiment, the identifying information can specify the identity of the speaker. For example, the identifying information can specify the speaker or call participant's name, an address, a contact telephone number, or any other identifying information, which may or may not correspond to the line from which the call participant is calling. In another embodiment, the identifying information can indicate whether the calling participant was successfully authenticated or verified.

**[0026]** More particularly, the identifying information can be sent to the subscriber as an encoded stream of digital information that is embedded within the digital voice stream. The IAS can identify which portions of the received audio signal are inaudible, for example using a psychoacoustic model. For instance, humans tend to have sensitive hearing between approximately 2 kHz and 4 kHz. The human voice occupies the frequency range of approximately 500 Hz to 2 kHz. As such, the IAS can remove portions of a voice signal, for example those portions below approximately 500 Hz and above approximately 2 kHz, without rendering the resulting voice signal unintelligible. This leaves sufficient bandwidth within a telephony signal within which the identifying information can be encoded and sent within the digital voice stream.

**[0027]** The IAS further can detect sounds that are effectively masked or made inaudible by other sounds. For example, the IAS can identify cases of auditory masking where portions of the voice signal are masked by other portions of the voice signal as a result of perceived loudness, and/or temporal masking where portions of the voice signal are masked due to the timing of sounds within the voice signal.

**[0028]** It should be appreciated that as determinations regarding which portions of a voice signal are inaudible are based upon a psychoacoustic model, some users will be able to detect a difference should those portions be removed from the voice signal. In any case, inaudible portions of a signal can include those portions of a voice signal as determined from the IAS that, if removed, will not render the voice signal unintelligible or prevent a listener from understanding the content of the voice signal. Accordingly, the various frequency ranges disclosed herein are offered as examples only and are not intended as limitations of the present invention.

**[0029]** The IAS can remove the identified portions, i.e. those identified as inaudible,

from the voice signal and add the identifying information in place of the removed portions of the voice signal. That is, the IAS replaces the inaudible portions of the voice signal with digital identifying information. As noted, the identifying information can include, but is not limited to, voice levels, stress levels, voice inflections, and/or emotional states as may be determined from a speaker's voice.

**[0030]** In step 230, the IAS sends the voice stream with the encoded and embedded identifying information to a receiving device of the subscriber. In step 235, the subscriber device receives the voice stream with the encoded identifying information and, in step 240, decodes the identifying information. In step 245, the receiving device can present the identifying information. For example, the identifying information can be presented visually or can be played audibly, for instance through a text-to-speech system. In step 250, the voice stream can be played audibly. In one embodiment of the present invention, the presentation of the identifying information and the playing of the voice stream can occur substantially simultaneously.

**[0031]** The inventive arrangements disclosed herein have been presented for purposes of illustration only. As such, neither the examples presented nor the ordering of the steps disclosed herein should be construed as a limitation of the present invention. For example, as noted, the present invention can be invoked prior to a telephone call or during a telephone call. The call participant also can be prompted for additional information as may be required to locate stored authentication data and perform identity verification.

**[0032]** The present invention can be realized in hardware, software, or a combination of hardware and software. The present invention can be realized in a centralized fashion in one computer system, or in a distributed fashion where different elements are spread across several interconnected computer systems. Any kind of computer system or other apparatus adapted for carrying out the methods described herein is suited. A typical combination of hardware and software can be a general purpose computer system with a computer program that, when being loaded and executed, controls the computer system such that it carries out the methods described herein.

**[0033]** The present invention also can be embedded in a computer program product, which comprises all the features enabling the implementation of the methods described

herein, and which when loaded in a computer system is able to carry out these methods. Computer program in the present context means any expression, in any language, code or notation, of a set of instructions intended to cause a system having an information processing capability to perform a particular function either directly or after either or both of the following: a) conversion to another language, code or notation; b) reproduction in a different material form.

**[0034]** This invention can be embodied in other forms without departing from the spirit or essential attributes thereof. Accordingly, reference should be made to the following claims, rather than to the foregoing specification, as indicating the scope of the invention.